

## Statistische Software (R)

Paul Fink, M.Sc.

Institut für Statistik  
Ludwig-Maximilians-Universität München

*Datenbeispiel*



Das Statistische Bundesamt veröffentlicht jedes Jahr die Zahl der Erwerbstätigen Inländer in Deutschland:

<https://www.destatis.de/DE/ZahlenFakten/Indikatoren/LangeReihen/Arbeitsmarkt/lrerw011.html>.

Die Daten können auch für die Jahre von 1970–2014 eingesehen werden. Als Variablen gibt es unter anderem die Anzahl Gesamtbevölkerung, der Arbeitnehmer, der Selbstständigen und Erwerbslosen (je in 1000).

Weil die Datenaufbereitung und das Einlesen im Fokus steht, sind die Daten bereits als Textdatei auf der Veranstaltungs-Homepage hinterlegt.

Paul Fink: Statistische Software (R) SoSe 2015

2

## Aufgaben

1. Laden Sie den (vereinfacht formatierten) Datensatz `erwerbstaetige.txt` von der Veranstaltungsseite ([http://www.statistik.lmu.de/institut/ag/agmg/lehre/2015\\_SoSe/StatSoft/material/erwerbstaetige.txt](http://www.statistik.lmu.de/institut/ag/agmg/lehre/2015_SoSe/StatSoft/material/erwerbstaetige.txt)) herunter und lesen Sie ihn als `erwerbstaetige` in R ein. Bringen Sie alle Variablen in ein sinnvolles Format.
2. Erstellen Sie 2 neue Variablen des Datensatzes (`erwerbstaetige` und `erwerbspersonen`), welche die Gesamtzahl der Erwerbstätigen (Arbeitnehmer und Selbstständige) sowie Erwerbspersonen (Erwerbstätige und Erwerbslose) je Jahr enthalten.

Tipp: `?transform`

## Aufgaben

3. Erstellen Sie eine weitere Variable, die den Anteil der Nicht-Erwerbspersonen an der Gesamtbevölkerung enthält. Lassen Sie sich eine statistische Kurzzusammenfassung dafür ausgeben.
4. Bestimmen Sie (nur mit R) die Jahre, in denen den Anteil der Nicht-Erwerbspersonen an der Gesamtbevölkerung am höchsten bzw. niedrigsten war.  
Tipp: Funktionen `which.min()` und `which.max()`

5. Visualisieren Sie ...

- (a) ... die empirische Dichtefunktion des Arbeitnehmeranteils an der Bevölkerung.
- (b) ... den zeitlichen Verlauf des Anteils der Nicht-Erwerbspersonen im Verlauf der Jahre.

6. Lesen Sie nun den (original formatierten) Datensatz

`erwerbstaetige_orig.txt` als `erwerbstaetige_orig` in R ein ([http://www.statistik.lmu.de/institut/ag/agmg/lehre/2015\\_SoSe/StatSoft/material/erwerbstaetige\\_orig.txt](http://www.statistik.lmu.de/institut/ag/agmg/lehre/2015_SoSe/StatSoft/material/erwerbstaetige_orig.txt)).

Führen Sie – soweit überhaupt notwendig – Änderungen an den in R eingelesenen Daten durch, damit sie im gleichen Format wie im data.frame `erwerbstaetige` sind.